J M S K
Jurnal Matematika, Statistika, & Komputasi
http:/journal.unhas.ac.id/index.php/jmsk
Vol.14, No. 2, 127-136, Januari 2018

p-ISSN: 1858-1382
e-ISSN: 2614-8811

Methods for Estimating Survival Time of Treatments for Renal Dialysis

Sri Astuti Thamrin[†]

Abstract

This papes discusses the theory and application of statistical methods for describing and analyzing survival times of the renal dialysis patients : a) from the first diagnosis until the time of death, and b) on each mode of given treatment. The paper also tries to predict the variables significantly effecting the survival time of renal dialysis patients. The paper makes use of and focuses on the data sets containing patient hospital records, patients' identity and hospital code centre. To meet the desired aims, the paper uses two prominent methods of survival analysis including the Kaplan-Meier and Cox Proportional Hazard model. The result shows that survival time on the first treatment depends on mode of treatment and it quite low approximately 18 days for median time on hospital outpatient CAPD. Similarly, survival time on the second treatment is quite low about 24 days for the median time on hospital outpatient CAPD. It was also indicated that the survival time of renal dialysis patient depends on the number of treatments, the number of treatment changes, place of treatment, age and the first treatment.

Keywords: Survival analysis, survival time, renal dialysis treatment.

1. Background

Chronic or unspecified renal failure was listed as a cause of death of 9160 Australians (7.1% of all deaths) (Australian Bureau of Statistics, special data request, 2002). Each year, more than 1700 people with end-stage renal disease (ESRD) start dialysis or receive a transplant (Rush, ANZDATA Registry Report 2001). Most would have had chronic renal impairment (CRI) for years. These figures suggest that the impact of CRI is substantial and that in order to prevent progression to ESRD, it is important to develop systems for its detection and management by providing relevant disease related information. Furthermore, it was reported that each year, there are about 2000 Australian adults who commence dialysis, half of whom are aged over 60 years.

Renal Dialysis is an "artificial method of maintaining the chemical balance of the blood when the kidneys failure happens." This method has increasingly developed in some countries. In NSW, Australia, the development of this renal dialysis is approximately 6% and this has exceeded the population growth of only 1% (Gibberd *et al.*, 2004).

In this paper, it will be applied a number of statistical methods to analyse renal dialysis data supplied by ANZDATA and NSW registry database in 1995 and has aims as follows:

- To describe and analyse the survival time of the renal dialysis patients: a) from the first diagnosis until the time of death, b) on each mode of given treatment and c) if they change the modes of given treatment.
- To predict the survival time for each treatment mode based on age, sex, region, time since diagnosis and the number of treatment changes.

[†] *Lecturer at Mathematics Department, Faculty of Mathematics & Natural Sciences, Hasanuddin University*

2. Overview of Used Methods

2.1. Survival Analysis

Historically, the trigger of the emergence of the use of survival analysis goes back to the World War II. During this war, the interest of scrutinizing the reliability or failure time of military equipment has seemingly inspired the new application of statistical methods, which then are known well as survival analysis methods. In the development of survival analysis techniques, even though extensively assigned in many areas of social, economic and in engineering for looking at reliability and failure time analysis, it has been primarily developed in the biological and medical researches. This is due to survival analysis methods being well suited for the research in the latter two sciences. As far as, for example clinical trials in medical research are concerned, the studies of intervention follow up could start without all experimental units enrolled at the beginning of study time and could finish before all experimental units had went through an event. In other words, there will be subjects of the experiment who choose to either quit participating or move away that is difficult to follow or die from some unrelated event (Smith *et al*., 1990).

To report survival times, some traditional methods of statistics such as the standard parametric and nonparametric statistics mainly describing the average survival could be used. The problem that could arise from applying those traditional methods is the censoring some objects of the study. Therefore, survival analysis needs to be based on censored observations. Censoring is what distinguishes survival analysis from other classical types of statistical analyses.

In general, survival time could be defined as "the analysis of data representing the time to occurrence of a certain point or endpoint (*time-to-event* data)" (Galbriath, 2004). The term "endpoints" could indicate death, release of pain and reappearance of symptoms. In most of cases, the application of survival analysis is aiming at identifying predictive variables for survival time or comparing the survival times of a number of different objects in a study.

To be applied, survival analysis requires at least three conditions including:

1. A well-defined time origin and therefore it is not ambiguous e.g. time of entry to study, which often corresponds to different times fir different individuals.
2. A scale for measuring the passage of time must be agreed, generally real time
3. A well-defined endpoint that implies that the meaning of failure must be completely clear (Galbriath, 2004; Cox & Oakes, 1984).

As mentioned before, censoring is very important in conducting survival analysis. Censoring takes place when an individual's life length, for example, is known to occur only in a certain period of time (Klein & Moeschberger, 1997).

There are three kinds of censoring. These include right censoring, left censoring and interval censoring. The terms of "right", "left", and "interval" censoring indicates the range in which survival time is known to locate. In relation to the censoring issues, this study mainly deals with right censoring.

In Galbraith's observation, right censoring is the common type of censoring in survival analysis and occurs when we only know that the survival time exceeds a certain value. To make it clear, she takes a five-year study of mortality from cancer as an example. According to her, survival times will be right censored for patients who:

- are alive at the end of the five year period
- drop out or become lost to follow-up during the study; or
- die from some other cause during the study (Galbriath, 2004; Klein & Moeschberger, 1997).

2.2. Kaplan Meier Survival Analysis

In dealing with censored data, the Kaplan Meier procedure is an important tool. It is a method for time to event models based on the estimation of conditional likelihood at each time point when a particular event takes place. The product limit of the probabilities estimates the survival rate at each point in time. By using this method, we can make comparisons of the entire survival (or failure) rates between two or more groups in order to see the effect of a particular mode of treatments on the survival time of the patients. The groups here could be age, sex, region etc.

The fundamental idea of Kaplan Meier model is to order the observation of the failure time variable and to let the right endpoint of I_i, τ_i, be the i^{th} ordered observation. For this reason τ_1, τ_2, ..., τ_k indicate distinct event times (Miller Jr., 1981). This can be illustrated as follows

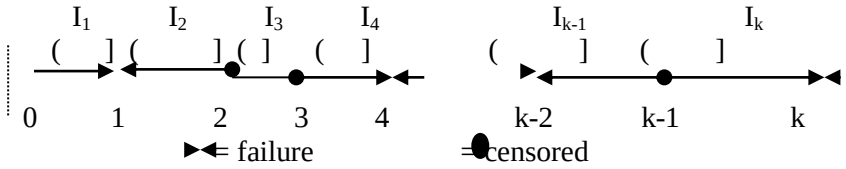


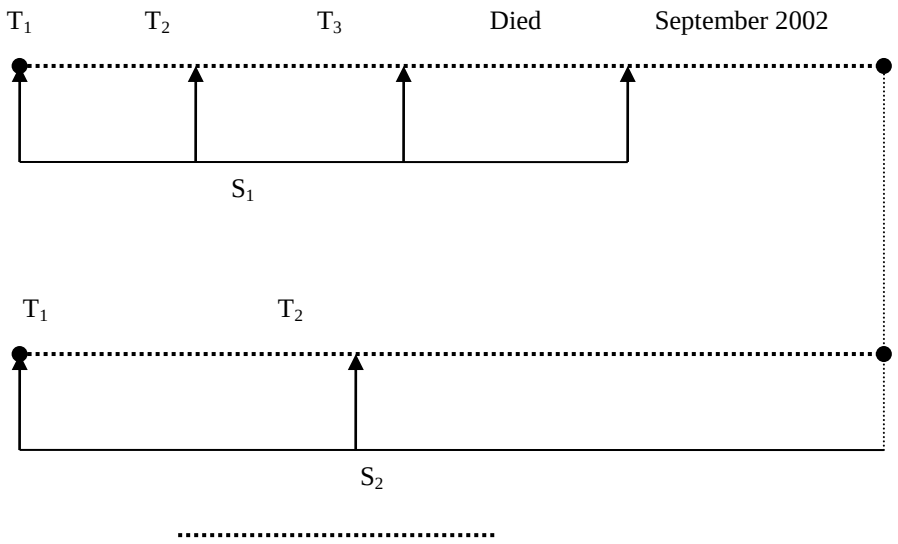
Figure 1. Ordering the Observations of The Failure Time Variable.

There are two kinds of observation that can be conducted including no tied observations and tied observations. The first observation considers that all the event times are different while the latter assume that uncensored observations take place just before the censored observations. Meanwhile the Kaplan Meier estimate of survival with no ties is

$$\hat{S}(t) = \prod_{X_i \le t} \hat{p}_i = \prod_{X_i \le t} \left(1 - \frac{1}{n_i}\right)^{\delta_{(i)}} = \prod_{X_i \le t} \left(1 - \frac{1}{n-i+1}\right)^{\delta_{(i)}} = \prod_{X_i \le t} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}$$

In calculating the Kaplan Meier survival estimates for one or more groups, we need a survival platform that can be used as a basis of getting information for more compound model. This Kaplan Meier Survival platform will indicate a plot illustrating the predicted survival function for each group and even for the whole sample. It also computes and lists survival function for each group and for the combined sample.

In this paper, the Kaplan Meier analysis is used to estimate three survival times of renal dialysis patients who have different treatment regimes. The first one is survival time on the first treatment, the second one is survival time on the second treatment and the last one survival time from the first diagnosis until either the time of death or the end of study. Illustratively this can be drawn as follows



Period of Study :
$T_{1,2,3}$: Time on a given treatment
S_1 : Time to death (uncensored)
S_2 : Time to death (censored)

Figure 2. The Three Survival Times of Renal Dialysis Patients.

Given the picture above, time on the first treatment is (T_2 - T_1) and time on the second treatment is ($T_3 - T_2$). Meanwhile time from the third mode of given treatment until the end of the study is considered as censored data.

After producing the Kaplan Meier estimate of the survival curve for renal dialysis patients, the median survival time was estimated. The median survival time is defined to be the smallest observed survival time for which the estimated survival function is less than 0.5 (Glantz, 2001).

2.3. Cox Proportional Hazard Model (Cphm)

The proportional hazard model (phm) is similar to common regression. The different feature of the phm compared with other regression models is that it can be used when we are dealing with data containing censored cases (Stephen, 2000).

In its simplest form, the phm can be the following equation:

$$h_i(t) = e^{x_i \beta} h_o(t)$$

$h_i(t)$ is the hazard at time t of the i^{th} individual and $h_o(t)$ is the baseline hazard at time *t*. X_i is a vector of covariate values corresponding to the i^{th} individual and β is a vector of coefficients to be estimated. Therefore, it is clear that $h_o(t)$ is dependent only on time and $e^{x_i \beta}$ depends only on the covariates and regression coefficients.

Given the simplest equation of the phm above, it can be furthermore noticed that if $X_i = 0$ then the hazard function of the i^{th} individual is the baseline hazard function. Additionally, if both sides of the equation are divided by $h_o(t)$, it can be seen that every value of the i^{th} individual's hazard function is s constant proportion of the baseline hazard. This is the reason why the term proportional is used.

Cox then proposed a semi-parametric method by which the regression parameters β can be estimated while leaving the baseline hazard function arbitrary (CCEB, 2004).

3. Results of Analysis of Renal Dialysis Data

3.1. Application for Kaplan Meier Model

In this section, the survival time of the renal dialysis patients: a) from the first diagnosis until the time of death, b) on each mode of given treatment are examined. In doing so, curves of survival are generated by using JMP 5.1.

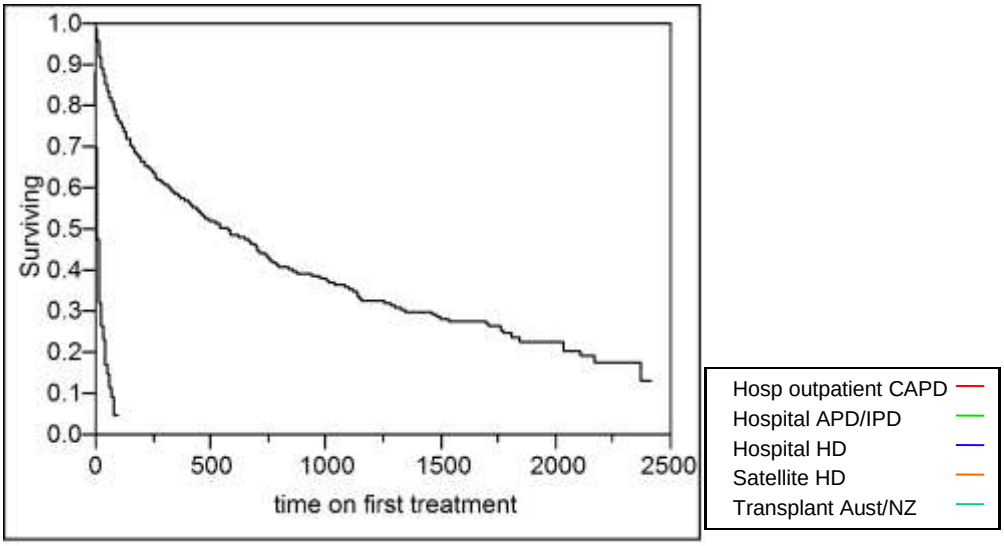


Figure 3. Survival Curve for The First Treatment By Treatment Mode.

Kaplan-Meier estimates are obtained for the first treatment. From the graph, it appears that survival is the best for transplant Aust/NZ than for others. After 152 days just over 1.52% of the transplant patients have changed treatment. However, hospital outpatient CAPD appears to have the shorter time before changing treatment. After 2 days, over 1.87% of these patients' have changed treatment. The p-value for the logrank test is 0.0000 with df = 4 and for the Wilcoxon test is <0.0001 with df = 4. We could conclude that there are statistically significant differences in time on the first treatment for all different types of the first treatment.

Overall, the number of patients who died from 1996 to September 2002 was 1257 and the number of patients who were still alive at the end of the data was 2729. All patients in the second modes of treatment, after an initial drop of approximately 0.35% in survival within the first days after first treatment, survival steadily decreases but not linearly, over the 2359-day period. The median survival time is 1334 days. Since the final time recorded is not a censored observation, the report does not indicate biased mean estimate. The mean estimate is not a lower bound for the true mean (1280 days).

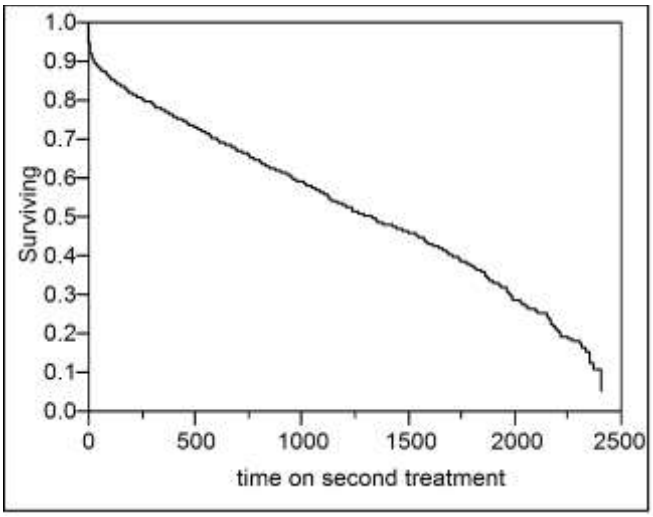


Figure 4. Survival Curve for The Second Treatment.

The smallest number of cases for the second treatment is hospital HD, whereas the large numbers of cases are home CAPD and satellite HD. However, the home CAPD has the larger variation than satellite HD.

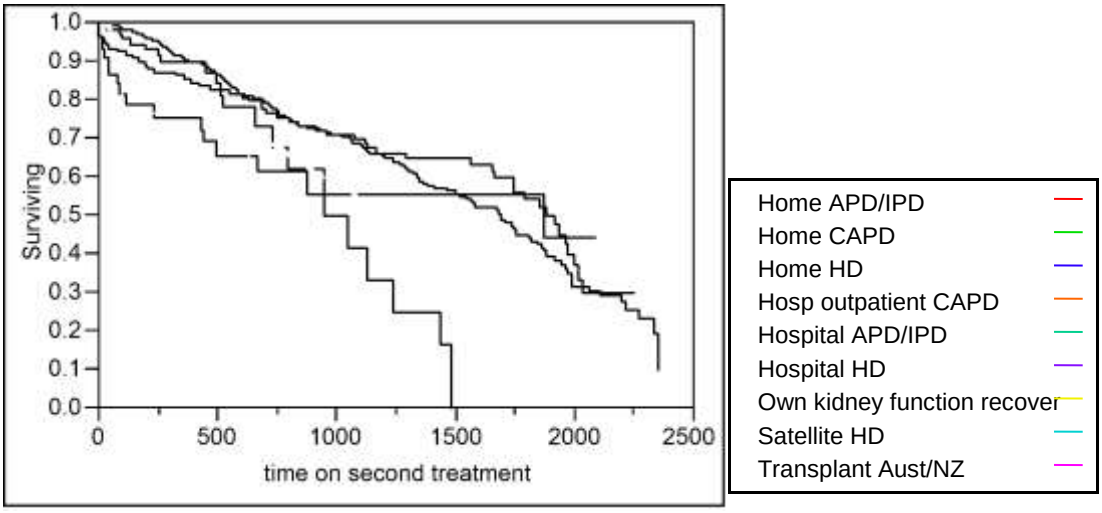


Figure 5. Survival Curve for The Second Treatment By Treatment Mode.

Kaplan-Meier estimates are obtained for the second treatment. From the graph, it appears that time on the second treatment is less for hospital outpatient CAPD than for others. After 3 days just over 6.27% of the hospital outpatient have changed treatment. The p-value for the logrank test and for the Wilcoxon test is <0.0001 with df = 8. We could therefore conclude that there are statistically significant differences in the first treatment for all different types of the second treatment.

The p-value for the logrank test is <0.0001 with df = 6, and for the Wilcoxon test is 0.05 with df = 6. We could therefore conclude that there are differences in time on the second treatment.

Tests Between Groups

Test	ChiSquare	DF	Prob>ChiSq
Log-Rank	3.8329	6	0.6993
Wilcoxon	11.6038	6	0.0714

The p-value for the logrank test is 0.69 with df = 6, and for the Wilcoxon test is 0.07 with df = 6. We could therefore conclude that there are no differences in time on the second treatment for home APD/IPD.

The time on treatment until death is given for those patients who started to receive the first mode of renal dialysis treatment in or after 1996 until death. We censored the patients who were still alive at the end of the data (September 2002), and were uncensored if the patient died.

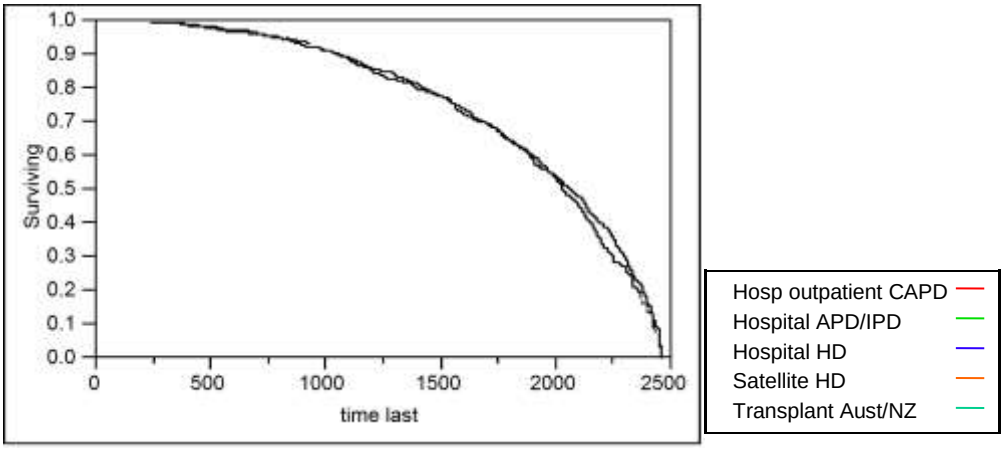


Figure 6. Survival Curve for All The First Diagnosis to Death

The Kaplan-Meier estimates are obtained for all types of the first mode of treatment to death. From the graph, it appears that compared to other treatments, survival time is the best for the transplant Aust/NZ and the worse for the three hospital treatments (HD, CAPD and APD/IPD). After 1864 days just over 6.25% of the transplant Aust/NZ patients have died. Meanwhile, the numbers of patients who have died under the hospital HD was 844 and after 2070 days just over 50.04% have died. The p-value for the logrank test and for the Wilcoxon test is <0.0001 with df = 4. We therefore conclude that there are statistically significant differences in survival time for the entire first mode of treatment till death.

Survival By Place of Treatment
- Hospital

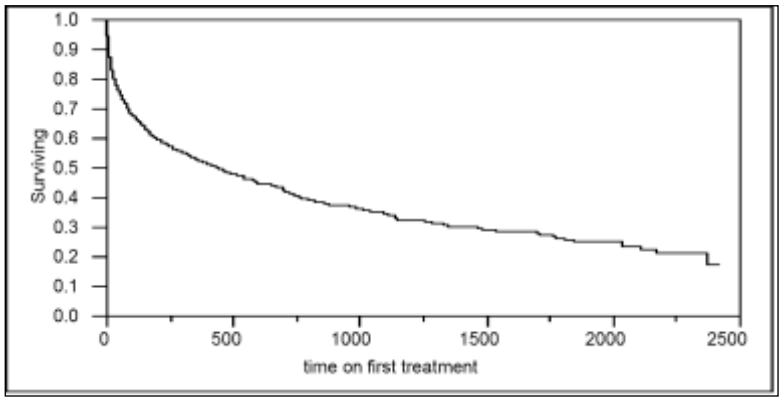


Figure 7. Survival Curve in Hospital for The First Modes of Treatment

Overall, the number of patients who died from 1996 to September 2002 was 1237 and the number of patients who were still alive at the end of the data was 2591. All patients in the first modes of treatment, after an initial drop of approximately 3.75% in survival within the first days after first treatment, survival steadily decreases but not linearly, over the 2353-day

period. The median survival time is 450 days. Since the final time recorded is a censored observation, the report indicates a biased mean estimate. The biased mean estimate is a lower bound for the true mean (894 days).

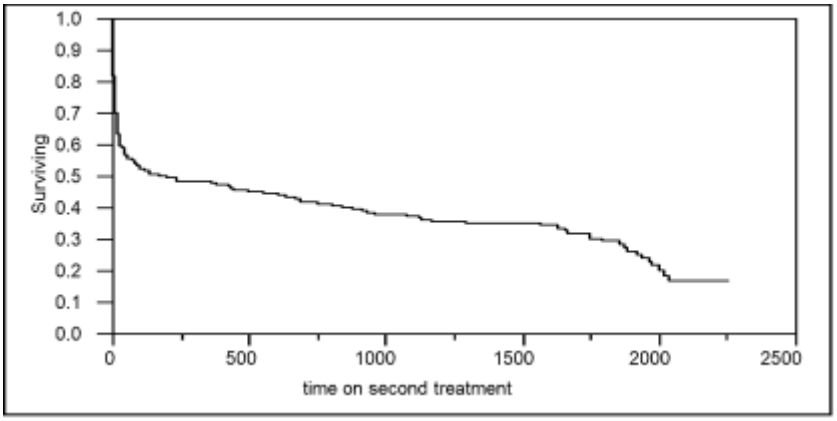


Figure 8. Survival Curve in Hospital for The Second Modes of Treatment.

Overall, the number of patients who died from 1996 to September 2002 was 413 and the number of patients who were still alive at the end of the data was 876. All patients in the second modes of treatment, after an initial drop of approximately 20.61% in survival within the first days after first treatment, survival steadily decreases but not linearly, over the 2023-day period. The median survival time is 202 days. Since the final time recorded is a censored observation, the report indicates a biased mean estimate. The biased mean estimate is a lower bound for the true mean (810 days).

- **Home**

No home-based on renal dialysis treatment was found at the modes of treatment.

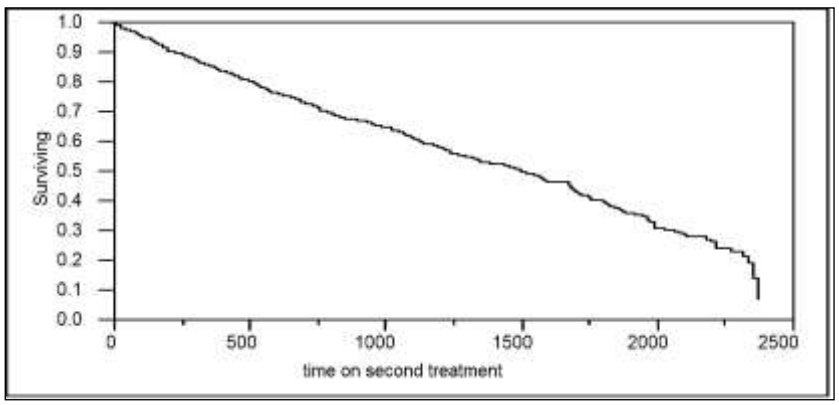


Figure 9. Survival Curve in Home for The Second Modes of Treatment.

Overall, the number of patients who died from 1996 to September 2002 was 472 and the number of patients who were still alive at the end of the data was 1104. All patients in the second modes of treatment, after an initial drop of approximately 0.25% in survival within the first days after first treatment, survival steadily decreases but not linearly, over the 2281-day period. The median survival time is 1505 days. Since the final time recorded is not a censored observation, the report does not indicate a biased mean estimate. The mean estimate is not a lower bound for the true mean (1400 days).

- **Satellite**

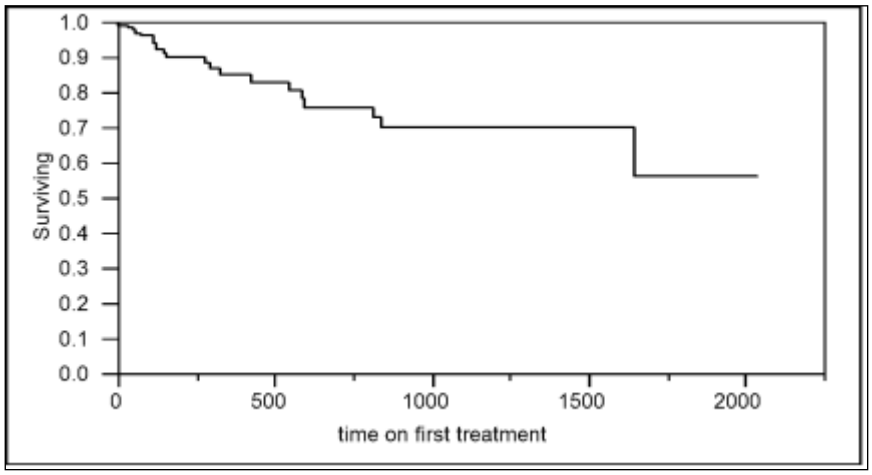


Figure 10. Survival Curve in Satellite for The First Modes of Treatment.

Overall, the number of patients who died from 1996 to September 2002 was 20 and the number of patients who were still alive at the end of the data was 138. All patients in the first modes of treatment, after an initial drop of approximately 0.25% in survival within the first days after first treatment, survival steadily decreases but not linearly, over the 2281-day period. Since the final time recorded is a censored observation, the report indicates a biased mean estimate. The biased mean estimate is a lower bound for the true mean (1283 days).

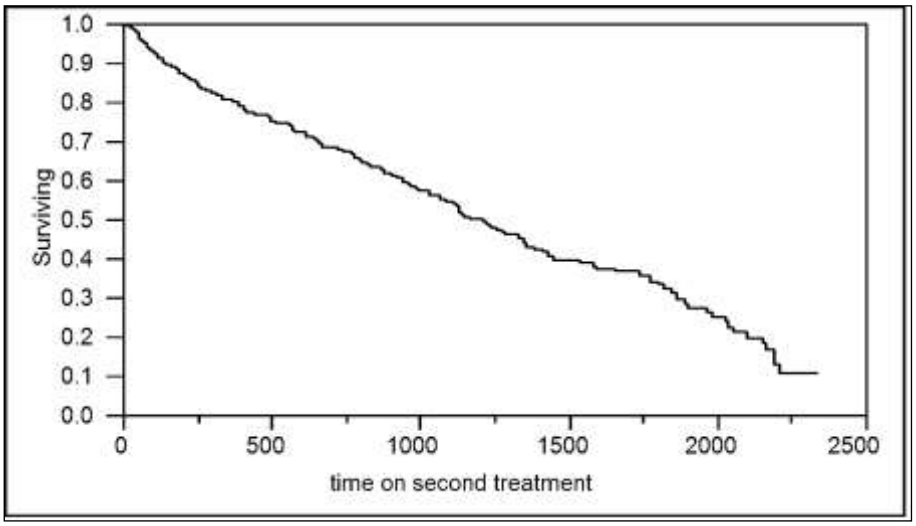


Figure 11. Survival Curve in Satellite for The Second Modes of Treatment.

Overall, the number of patients who died from 1996 to September 2002 was 241 and the number of patients who were still alive at the end of the data was 527. All patients in the second modes of treatment, after an initial drop of approximately 0.13% in survival within the first days after first treatment, survival steadily decreases but not linearly, over the 2201-day period. The median survival time is 1220 days. Since the final time recorded is a censored observation, the report indicates a biased mean estimate. The biased mean estimate is a lower bound for the true mean (1225 days).

3.2. Application for Cox Proportional Hazard Model (Cphm)

In this section, the identification of some variables that affecting survival time of renal dialysis patients is conducted by using STATA 8. Initially, a frequency check of all variables was undertaken. The variables of interest for these analyses are: sex, number of treatments used (treat_used), age1, and number of treatment changes (treat_change), post code (postcd), year and first treatment (treat1_cd). Backward stepwise regression was undertaken (see appendix d). The final model included 5 variables significant at the 5% level. Those variables embrace: age (category2: 5-14 years, category 3: 15-24 years, category 4: 25-34 years, category 5: 35-44 years, category 6: 45-54 years and category 7: 55-64 years), number of change treatments, year (1997 to 2001) and the first treatments (satellite HD and transplant Aust/NZ).

4. Conclusion

Having analysing the survival time of renal dialysis patients who have different treatment regimes it is clearly can be seen that the hospital HD treatment is the command mode of the first treatment, which is increased every year. It was also found that in 1996, 25% of patients had changed treatment after 50 days of treatment but it was 307 days in 2001. Meanwhile, home CAPD is the comment mode of the second treatment used by renal dialysis's patients. Overall survival time of renal dialysis patients depend on the number of treatments, the number of treatment changes, places of treatment, age and the first treatment. Specifically we can identify the most significant variables that affect the survival time of the renal dialysis patients. These variables consists of age group of 15-24 years old and 75-84 years old, number of treatment used, number of treatment changes, satellite for the place of treatment and transplant AUST/NZ for the first treatment.

References

[1] CCEB, 2004, "Parametric and semiparametric methods for survival data", *Module 6 of Advanced Biostatistics*, University of Newcastle, Australia.

[2]. Cox, R.D., 1984, "*Analysis of Survival Data*", Chapman and Hall, Melbourne, Australia.

[3]. Efron, B., 1988, "Logistic regression, survival analysis and the Kaplan-Meier curve", *Journal of the American Statistical Association*, Vol. 83, No. 402.

[4]. Galbriath, S., 2004, "Statistics for Biomedical research", *Course Notes*, University of New South Wales, Australia.

[5]. Glant, S., 2001, "*Primer of Biostatistics*", McGraw-Hill, USA.